

**METHOD AND APPARATUS FOR AUDIO/IMAGE SPEAKER  
DETECTION AND LOCATOR**

**Abstract of the Disclosure**

A method and apparatus for a video conferencing system using an array of two  
5 microphones and a stationary camera to automatically locate a speaker and electronically  
manipulate the video image to produce the effect of a movable pan tilt zoom ("PTZ") camera.  
Computer vision algorithms are used to detect, locate, and track people in the field of view of a  
wide-angle, stationary camera. The estimated acoustic delay obtained from a microphone array,  
consisting of only two horizontally spaced microphones, is used to select the person speaking.  
This system can also detect any possible ambiguities, in which case, it can respond in a fail-safe  
way, for example, it can zoom out to include all the speakers located at the same horizontal  
position.